

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): Karoline Freeman, Ala Szczepura, Leeza Osipenko

Article Title: Non-invasive fetal RHD genotyping tests: A systematic review of the quality of reporting of diagnostic accuracy in published studies

Year of publication: 2008

Link to published version: <http://dx.doi.org/10.1016/j.ejogrb.2008.10.010>

Publisher statement: None



Contents lists available at ScienceDirect

## European Journal of Obstetrics & Gynecology and Reproductive Biology

journal homepage: [www.elsevier.com/locate/ejogrb](http://www.elsevier.com/locate/ejogrb)



### Review

# Non-invasive fetal *RHD* genotyping tests: A systematic review of the quality of reporting of diagnostic accuracy in published studies

Karoline Freeman<sup>\*</sup>, Ala Szczepura, Leeza Osipenko

Warwick Medical School, Warwick University, Gibbet Hill Road, Coventry CV4 7AL, UK

#### ARTICLE INFO

##### Article history:

Received 19 November 2007  
Received in revised form 8 October 2008  
Accepted 17 October 2008

##### Keywords:

RhD  
Non-invasive prenatal diagnosis  
Test accuracy  
Quality assessment  
Sensitivity and specificity

#### ABSTRACT

Articles reporting the diagnostic accuracy of non-invasive prenatal diagnostic (NIPD) tests for *RHD* genotyping using fetal material extracted from maternal blood have been published steadily for over a decade. Health care providers in Europe have started to use this technology for management of the small number of sensitised pregnancies (ca. 220–600 per annum in the Netherlands, Germany, France and the UK). Scientists and clinicians are also advocating widespread implementation for the far larger number of non-sensitised RhD-negative pregnancies (ca. 34,000–125,000 per annum in the same countries). Large-scale, prospective trials are only now underway. Estimates of the technical performance of these tests are currently based on results from small-scale studies, together with formal meta-analysis. The issue of early assessment of test performance is one faced by many new genetic tests. As part of a wider study we have investigated the quality of reporting of diagnostic accuracy in publications and produced guidelines for future studies.

A systematic search of the literature identified 27 papers which met predefined inclusion criteria. All 27 papers were, first, assessed against an international quality (STARD) checklist for reporting of diagnostic accuracy and, second, against our own in-house NIPD proforma to assess the implications of the quality of reporting specifically for the RhD NIPD test. Authors were found to generally present an optimistic view of NIPD, bearing in mind weaknesses identified in reporting and conduct of their studies and the analysis of results, as evidenced by the low STARD scores. The NIPD proforma identified that specific biases were potentially introduced through selective population sampling and/or failure to report the make-up of the population tested, omission of inconclusive results, inconsistencies in the handling of repeat results on a sample, and lack of adequate controls. These factors would inevitably affect the validity of diagnostic accuracy as reported in individual publications, as well as any subsequent meta-analyses. Together, published reports to date may provide a biased picture of the actual potential of NIPD testing for fetal *RHD* genotyping. Generalisation of the available evidence on diagnostic accuracy, especially to large-scale implementation of NIPD testing of non-sensitised women, will also require that decision makers consider further aspects such as test reliability and cost of routine testing in clinical practice.

It is recommended that all studies of diagnostic accuracy of NIPD tests adhere to the STARD quality checklist in order to improve reporting, thereby, minimising bias and increasing the comparability of studies. Researchers should also consider specific shortcomings for NIPD and avoid selective participant sampling; report population characteristics; report handling of replicate sampling as well as their failure rates; and include controls for genotypes tested in the study. Furthermore, meta-analyses should consider the quality, as well as the sample size, of NIPD studies in their analysis. Larger trials, required to produce results that are valid and meaningful for clinical practice, must also adhere to these reporting standards.

© 2008 Published by Elsevier Ireland Ltd.

<sup>\*</sup> Corresponding author. Tel.: +44 2476 574870; fax: +44 2476 524963.

E-mail addresses: [k.freeman@warwick.ac.uk](mailto:k.freeman@warwick.ac.uk), [Ala.Szczepura@warwick.ac.uk](mailto:Ala.Szczepura@warwick.ac.uk), [J.H.Hyde@warwick.ac.uk](mailto:J.H.Hyde@warwick.ac.uk) (K. Freeman).

## Contents

1.	Introduction .....	000
2.	Methods .....	000
2.1.	Literature search strategy .....	000
2.2.	Selection process for papers for review. ....	000
2.3.	Data extraction and quality assessment .....	000
3.	Results .....	000
3.1.	Assessment against the STARD checklist. ....	000
3.2.	Assessment against NIPD proforma. ....	000
3.3.	Study population .....	000
3.4.	Study conduct and data analysis. ....	000
3.5.	Study design .....	000
4.	Discussion .....	000
	Acknowledgements .....	000
	References .....	000

## 1. Introduction

Fetal *RHD* genotyping from maternal whole blood [1], fetal cells [2] or cell free fetal DNA [3] in maternal blood has been feasible for over a decade. Following the first papers, research interest in this form of non-invasive prenatal diagnosis (NIPD) has become global. Publications in Europe, America and Australia aim to improve the management of pregnancies of women who are rhesus negative [2,4–9]. In white Caucasian populations about 10% of all pregnancies involve an RhD-negative mother and an RhD-positive fetus, potentially placing the mother at risk of sensitisation and future babies at risk of haemolytic disease of the fetus and newborn (HDFN).

Anti-D immunoglobulin can be given to prevent a woman producing antibodies against fetal RhD-positive blood cells. Introduction of postnatal anti-D prophylaxis (within 72 h of delivery) in the late 1960s reduced sensitisation and HDFN rates considerably internationally. Combined with antenatal prophylaxis for high-risk events, such as amniocentesis, this form of targeted prophylaxis has proved highly effective and cost-effective [10]. Consequently, the number of sensitised RhD-negative women who give birth every year is lower than 1% of total births.

For non-sensitised women, routine antenatal anti-D prophylaxis (RAADP) at 28–30 weeks gestation was introduced in the mid-1990s, and is now available in most European countries. However, 40% of RhD-negative women receive this blood product unnecessarily because the fetus is not RhD-positive. Fetal *RHD* genotyping has the potential to enable targeted antenatal prophylaxis only for women whose pregnancy is at risk. This might address any issues of limited anti-D supplies and perceived risk associated with unnecessary administration of a blood product.

The NIPD *RHD* genotyping test has already started to have an impact on the management of sensitised women, in countries where this technology is available, by replacing invasive procedures like amniocentesis for the determination of fetal RhD status. Replacement of amniocentesis is desirable since the procedure can promote feto–maternal blood exchange as well as being associated with a small increased risk of miscarriage [11,12].

The current state of development of the RhD NIPD test is promising. Several studies have reported high accuracy rates [13–16] and these have prompted the initiation of larger scale trials in the Netherlands, Germany and the UK [17–18]. Reports in the literature have already led to initial use of the test for management of sensitised women. An increasing focus of debate is now on widespread implementation of the test in non-sensitised pregnancies [19]. Discussion of markers (exons and/or introns) to be used for *RHD* genotyping, and the applicability of NIPD tests for

routine clinical use in diverse populations form part of this debate [13,17].

A recent meta-analysis undertaken by Geifman-Holtzman et al. has reported sensitivity and specificity values of 0.986 and 0.954 respectively, and an overall positive predictive value of 0.990 and negative predictive value of 0.921 [20]. Some studies were excluded from this meta-analysis on the basis of sample numbers, but aspects such as study quality were not considered. Because assessment of the quality of reporting of diagnostic accuracy can be problematic, a world-wide Delphi panel has identified agreed international quality criteria for assessing the reporting of such studies [21]. The resulting checklist (STARD –standards for reporting studies of diagnostic accuracy) is now required from all authors of health technology assessment reports of diagnostic technologies in the UK, and increasingly requested by editors of international journals. We have undertaken an independent systematic literature search and assessed the quality of reporting of all identified articles using the STARD checklist. We further investigated the implications of any shortcomings on the generalisability of NIPD study results, in order to identify key aspects influencing test reliability and any underlying reporting biases. An NIPD assessment proforma has been produced as well as recommendations that are specific and relevant for reporting diagnostic accuracy of RhD NIPD tests.

## 2. Methods

### 2.1. Literature search strategy

Published articles were identified by systematic searches of electronic databases from 1966 until January 2007; these included PubMed, Ovid Medline, Ovid Embase, the Cochrane Library, the National Library for Health (UK), Online Computer Library Center (OCLC) and the Conference Papers Index. Text words and MeSH headings used separately and in combination included: prenatal diagnosis, Rh, fetal cells, fetal DNA, maternal blood, serum, plasma, Rh alloimmunisation. Bibliographies of all papers identified were examined. Searches for related articles by topic and author were carried out in PubMed where possible. No restriction was set on publication type or language.

### 2.2. Selection process for papers for review

All retrieved publications were read and assessed in respect of study type, patient/sample numbers and the stated study aim/research question. Papers not in English were translated. Studies that contained information on the diagnostic accuracy of RhD NIPD were identified. Short abstracts, letters and case reports were

**Table 1**

Coverage of the STARD checklist items by 27 papers reporting the diagnostic accuracy of RhD NIPD.

Section and topic	No.	Checklist items	Coverage <sup>a</sup> (%)
Title/abstract/ keywords	1	Identify the article as a study on diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	61 <sup>b</sup>
Introduction	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	92 <sup>c</sup>
Methods			
Participants	3	Describe the study population: the inclusion and exclusion criteria, setting and locations where the data were collected.	32
	4	Describe participant recruitment: was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	78
	5	Describe participant sampling: was the study population a consecutive series of participants defined by selection criteria in items 3 and 4? If not, specify how participants were further selected.	12
	6	Describe data collection: was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	34
Test method	7	Describe the reference standard and its rationale.	54
	8	Describe technical specification of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	98
	9	Describe definition and rationale for the units, cutoffs and/or categories of the results of the index test(s) and the reference standard.	0
	10	Describe the number, training and expertise of the persons executing and reading the index tests and the reference standard.	4
	11	Describe whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	28
Statistical methods	12	Describe methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	8
	13	Describe methods for calculating test reproducibility, if done.	0
Results			
Participants	14	Report when study was done, including beginning and ending dates of recruitment.	8
	15	Report clinical and demographic characteristics of the study population (e.g. age, sex, spectrum of presenting symptoms, comorbidity, current treatments and recruitment centers).	68
	16	Report the number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended).	16
Test results	17	Report time interval from the index tests to the reference standard, and any treatment administered between.	48
	18	Report distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	0
	19	Report a cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	88
	20	Report any adverse events of index tests and reference standard.	4
	21	Report estimates of diagnostic accuracy and measures of statistical certainty (e.g. 95% confidence intervals).	12
	22	Report how indeterminate results, missing responses and outliers of index tests were handled.	62
	23	Report estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	0
	24	Report estimates of test reproducibility, if done.	0
Discussion			
Discussion	25	Discuss the clinical applicability of the study findings.	98

<sup>a</sup> Coverage is a measure of how well the studies have adhered to items of STARD checklist (coverage =  $\sum(\text{scores})/\text{total papers}$ ), where studies covering checklist item well score 1 point, partially 1/2 point and no coverage 0 points).

<sup>b</sup> Phrases used to identify studies as diagnostic papers included: accuracy/detection rate, sensitivity and specificity, evaluation of test reliability, efficiency and feasibility.

<sup>c</sup> Studies scored as long as they mentioned: evaluation or determination of test accuracy, evaluation of test reliability, efficiency and feasibility of test.

excluded due to limited data content. Similarly, small studies with fewer than 10 samples were excluded. Phrases used to identify studies as diagnostic accuracy papers included: accuracy/detection rate; sensitivity and specificity; evaluation of test reliability; efficiency and feasibility.

### 2.3. Data extraction and quality assessment

A spreadsheet was prepared to extract information from the articles finally selected. This included: year and country of publication; number of patients/samples; source and type of fetal material; number of replicates and exons tested; number of positive replicates required for a result to be categorised as positive; test failure/inconclusive result rates; accuracy measures reported.

Papers were then scored against the STARD checklist of 25 items, resulting in a score out of 25 for each paper which

corresponded to the paper's quality as a study reporting diagnostic accuracy. Coverage of the 25 STARD items by the 27 studies was also reported (Table 1).

An in-house proforma was developed by collating the eight main points of the data extraction form and categorising these under the key elements of design deficiencies [22]. Papers were also assessed against this.

### 3. Results

The literature search resulted in the identification of 51 publications. The selection procedure described above excluded 24 papers (Fig. 1) leaving 27 studies for final quality assessment (Table 2). Of these, 25 were in English, 1 in French and 1 in Polish. Publication dates ranged from 1996 to the end of 2006. Findings were published in a wide range of journals and originated from 15 countries.

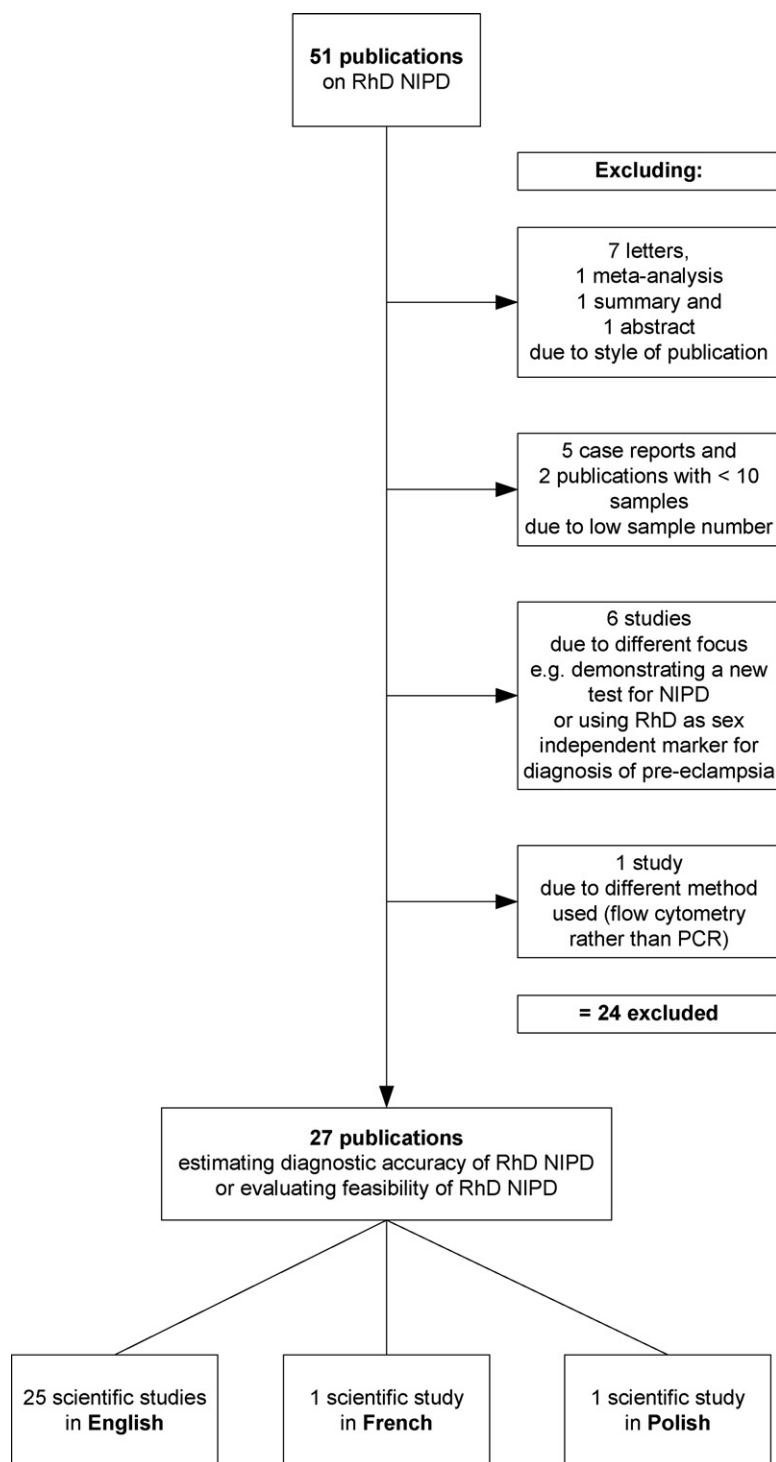


Fig. 1. Flow chart describing the selection process of identified literature with reasons for exclusion.

### 3.1. Assessment against the STARD checklist

All studies were assigned a low score based on the STARD checklist. Scores ranged from as low as 5/25 to 13/25 points. Studies demonstrated weaknesses in particular elements of the STARD checklist, as shown in Table 1. Quality of reporting was not related to publication date or country of study. The great majority of authors identified their paper as a report on diagnostic accuracy (STARD item No. 1) and nearly all authors stated their aim to be assessment of the feasibility of RhD NIPD tests (No. 2). However,

fewer than one in three actually used the recommended MeSH heading 'sensitivity and specificity' required for indexing papers on diagnostic accuracy.

When reporting study methods, articles generally lacked information on the population providing the samples, including how participants were recruited and which women were selected, and whether data was collected prospectively or retrospectively (Nos. 3–6). All these will affect the comparability as well as quality of study findings. Linked to this, the degree of blinding of test readers (No. 11) and the statistical methods used to quantify

**Table 2**

Weaknesses affecting the generalisability of reported accuracy rates of RhD NIPD in 27 studies.

Study	Study population		Study conduct		Data analysis		Study design		Refs.
	Selective participant sampling	Lack of reporting ethnicity and/or sensitisation status of participants	Lack of reporting no. of replicates (if done) used for overall study outcome	Lack of reporting failure rate	Lack of inclusion of reported failure rate into analysis	Difference in reported and adjusted accuracy	Lack of control for presence of fetal DNA	Lack of known genotypes in study as control	
1					✓	✓	✓		[13]
2		✓			✓	✓		✓	[14]
3		✓			✓	✓	✓	✓	[35]
4	✓				✓	✓	✓		[15]
5	✓	✓		✓	✓		✓	✓	[26]
6					✓	✓	✓	✓	[30]
7		✓		✓	✓		✓	✓	[33]
8				✓	✓	✓	✓	✓	[28]
9							SRY only		[32]
10		✓						✓	[31]
11		✓			✓	✓		✓	[16]
12	✓	✓			✓	✓	✓		[36]
13		✓						✓	[25]
14		✓		✓	✓		✓	✓	[34]
15		✓	✓				✓		[4]
16		✓					✓		[39]
17	✓	✓	✓		✓	✓	✓		[5]
18	✓	✓	✓	✓	✓		SRY only	✓	[9]
19		✓						✓	[6]
20		✓					✓		[37]
21	✓	✓					✓	✓	[23]
22		✓		✓	✓	✓	✓	✓	[7]
23		✓			✓	✓	✓		[29]
24	✓	✓		✓	✓		✓	✓	[8]
25		✓	✓	✓	✓		✓		[38]
26	✓	✓		✓	✓		✓	✓	[24]
27		✓		✓	✓	✓	✓	✓	[2]

uncertainty (No. 12) were also rarely reported in studies, while test reproducibility (No. 13) was not covered at all. On the other hand, materials and methods (No. 8) were well described. The reference standard (No. 7) was mentioned in all but two studies but rarely described comprehensively. Numbers and training of test readers (No. 10) were rarely reported, and cut-off or category of result (No. 9) was not covered in any of the studies.

When reporting results, the actual date of the study was rarely recorded (No. 14). Details of the study population generally only included information on the gestational age of the fetus but not the characteristics of women (No. 15). The proportion of participants satisfying criteria for inclusion was also rarely reported (No. 16). Reporting of test results was often vague with limited details of diagnostic test parameters or measures of statistical certainty (No. 21), while estimates of variability between different patient populations, test readers or centres together with estimates of test reproducibility (Nos. 23 and 24) were not reported in any study. Handling of missing responses and indeterminate results (No. 22) was reported in two-thirds of papers. Adverse events were not reported (No. 20) although the relevance of this item might be questioned, as might that relating to the severity of disease in those tested (No. 18). The time interval between index test and reference standard could only be estimated from the type of reference standard used, i.e. postnatal serology, amniocentesis or CVS (No. 17). A cross tabulation of results was presented in nearly all studies (No. 19). Finally, discussion of the clinical applicability of the study findings was covered in all studies (No. 25).

A detailed comparison of STARD checklist scores across papers would appear to indicate that articles have consistent weaknesses in their reporting, and published studies may, therefore, not be free from bias, with some indication that reported diagnostic accuracy rates may be overestimates. In terms of their applicability to the

service setting, certain STARD items could have major implications for clinical implementation of NIPD testing for fetal *RHD* genotyping, while others might not apply directly to the NIPD test being evaluated. For instance, a description of cut-offs and categories (STARD list item No. 9) is not obviously applicable to RhD NIPD. Any differential implication that specific items might have for clinical application of RhD NIPD is not reflected in the final score since all STARD items carry equal weight. The STARD score, therefore, should not be interpreted as a stand-alone measure of the quality of studies of RhD NIPD diagnostic accuracy. There may be specific questions and unique characterisations which are important for this type of genetic diagnostic tool.

### 3.2. Assessment against NIPD proforma

An in-house proforma translated the list of general shortcomings identified by STARD into areas of weakness specifically relevant to RhD NIPD accuracy. Four main areas, previously reported by the STARD group [22], incorporating eight specific aspects, were identified as most likely to compromise the internal as well as external validity of NIPD studies (Table 2). A lack of discussion of flaws in these areas might lead to inappropriate decisions on adoption of the test.

### 3.3. Study population

Table 2 clearly indicates that a number of researchers concentrated on samples from at risk pregnancies (e.g. Down syndrome), including women who may not resemble the general RhD-negative pregnant population [5,8,9,15,23–26]. Since recruitment was carried out selectively, test accuracy reported in such studies might be an overestimate because certain fetal conditions



such as Down syndrome can lead to higher levels of fetal DNA in the maternal circulation [27], consequently increasing the sensitivity of the test. Only 5 studies made reference to the sensitisation status of the sample population [6,15,16,28,29]. Poor characterisation of the sample population or selective sampling might affect the reliability of reported results equally. Selective sampling was taken even further in two studies that only recruited RhD-negative pregnant women with RhD-positive fetuses [15,29]. Both studies excluded any possibility of producing false positives and did not sufficiently challenge the test because a certain test outcome was anticipated. This is also true for one study that included samples from four women with weak D phenotypes [30]. Fetuses of women with such a phenotype cannot currently be tested non-invasively. While identification of such individuals is an important feature of the test in order to avoid false positives, knowledge of inclusion of such samples, again, implies an anticipation of identifying these four women. Restriction of samples to Caucasians also reduced the validity of findings, as did failure to report the ethnic make-up of the study population (20 studies). This is particularly relevant for *RHD* genotyping because of the genetic variation observed in different ethnic groups. Only four studies specifically identified their study populations as of mixed ethnicity [13,25,31,32].

### 3.4. Study conduct and data analysis

A third aspect concerns the handling of data on replicate samples. Twenty studies reported the number of replicates used, which ranged from two to seven for fetal DNA samples and up to 11 when using fetal cells. Out of these, five studies used a different number of replicates for different samples [9,14,30,33,34] and four studies presented no, or insufficient, information on how replicates were handled. Seven studies either did not include replicates or did not report them. Protocols for conversion of replicate outcomes into an overall test result differed significantly across the studies, ranging from a requirement for one or more positive replicate outcome [2,13,28,29,30,32–34] to two or more [6,16,25,35] or all replicates [14,15,31,35,36] to be positive for an overall positive result.

A fourth aspect relates to reporting of test failure rates and whether the reported accuracy is adjusted for the number of retests. A test failure rate, which is derived from the number of excluded samples due to inconclusive results, lack of material or maternal variant genes, was reported in only 12 studies [4–6,13–16,25,29–31,35]. Out of these, only one group of authors considered the failure rate in their analysis of test accuracy [31] and one mentioned the revised overall success rate only in their comments [15]. In all studies that provided data on their test failure rate, inclusion of these samples would result in a lower accuracy [13–16,25,29–31,35]. This was also demonstrated in a recent meta-analysis [20].

### 3.5. Study design

A final two aspects concern the controls included in a study. Ideally, for RhD NIPD these should include (a) controls of known RhD genotypes and (b) controls for the presence of fetal material. Use of control samples of known RhD type (positive as well as negative) and run at the same time as the index test was reported in only 10 studies [4,5,13,15,25,29,32,37–39]. Only two of these were controls from pregnant women with known RhD-positive or -negative fetuses. The other studies used control samples from males, non-pregnant females or combinations of both. Controls of known RhD variants that some studies claimed to be able to distinguish were not reported in those papers.

Use of a control for the presence of fetal DNA was reported in only seven studies [9,13,14,16,31,32,36]. There was no consistency in the type of control used. Y-linked markers were used in two studies but these are unable to provide a positive control for female fetuses [9,34]. Insertion deletion polymorphisms (IDPs) were used in four studies; these are sex independent and appear reasonably robust [14,16,31,36]. However, using 11 [14] and 8 [31] IDPs, the authors report that the presence of fetal DNA could not be confirmed in a small number of samples. Spiking samples with mouse DNA was not sufficient as a control for the presence of DNA [5], since a true control requires DNA of fetal origin. Employment of *RHCE* as a control for fetal DNA in a further study was not sufficient because of the maternal genetic background [13].

## 4. Discussion

Articles which report the diagnostic accuracy of new non-invasive prenatal diagnostic tests for *RHD* genotyping are difficult to identify through systematic searches of the literature. Indexing of the articles identified was found to be poor, as has been reported by other diagnostic test reviews [40]. Of those finally identified, less than one-third were found to have included the correct MeSH headings for diagnostic accuracy.

A detailed appraisal of all retrieved articles has identified a generally poor quality of reporting when papers are scored against international standards for studies reporting diagnostic accuracy and against an in-house NIPD proforma. The weaknesses identified are similar to those mentioned in two recent papers [41,42]. The shortcomings observed will be significant if they cause the final test accuracy, after adjustment, to fall below a level that is acceptable in the context of their proposed clinical use [41].

A recent publication reviewing the current stage of RhD NIPD test development reported a test diagnostic accuracy of >99% [43]. At the same time, a meta-analysis of published studies has reported an accuracy of 94.8% [20]. While an accuracy of >99% might be sufficiently high for a diagnostic test in clinical practice, an accuracy rate of 94% has been identified as suboptimal for clinical practice [41]. Further in their meta-analysis Geifman-Holtzman et al. [20] report an overall accuracy rate of only 91.4% when considering all samples, including those that had been excluded by study authors. Our review indicates widespread exclusion of inconclusive results, and of samples with insufficient fetal material or those with maternal variant genes, which will both have important implications for population wide implementation. This is apparent when one considers that in clinical practice all such samples excluded from an analysis would represent individuals for whom the test cannot produce a result. Although the two main factors that will limit broad-scale application of RhD NIPD are false negatives and false positives, it is, therefore, important to also achieve a reduction in the number of excluded samples.

False negatives may be due to lack of fetal material as well as suboptimal test sensitivity. However, an internal control to routinely confirm the presence of fetal DNA in the sample is not yet available [13]. Sex independent controls such as IDPs cannot be used as internal controls and highly polymorphic short tandem repeats (STRs), although discussed as a potential control, are reported to exhibit much lower sensitivity than the sensitivity for *RHD* genotyping [25]. Lack of a suitable control is currently a major limitation of the test for clinical practice.

Our review also identified particular study weaknesses in patient sampling. Selective participant sampling can result in a study population that does not resemble the general population to be tested in clinical practice. Similarly, a lack of consistency in the handling and reporting of replicates will limit the generalisability

### Recommendations for the reporting of genetic diagnostic test accuracy for RhD NIPD

#### Replicates / samples/ patients:

- Report clearly the number of patients, samples and replicates in the study
- Report outcomes for all patients, samples and replicates
- Consider implication of number of replicates employed (note: concentration of DNA is major limitation of NIPD, therefore, a great number of replicates is not desirable)
- Report how many positive replicates are considered for overall positive outcome

#### Repetition of test:

- Report the number of repeat tests needed per sample

#### Controls:

- Include controls of known RhD genotypes, i.e. RhD+, RhD- and variants included in testing strategy
- Consider the need of sex independent, internal controls for the presence of
- Record results of all control outcomes

#### Patient sampling:

- Report patient sampling procedure (consecutive patient series, prospective/retrospective study)
- Report dates and centres of recruitment and factors that recruitment was based on

#### Results:

- Report true positives, true negatives, false positives, false negatives

#### Failure rate:

- Report failure rate: number of inconclusive and excluded samples / total number of samples

#### Accuracy rate:

- Report the overall accuracy rate including all failed samples, excluded patients / samples and inconclusive results

#### Discussion:

- Discuss in terms of overall test accuracy and applicability to clinical practice:
  - implication of using number of samples/patients in analysis if applicable
  - implication of number of replicates considered in overall test outcome
  - implication of repeat tests
  - implications of selective patient sampling
  - the overall accuracy rate including failed samples, excluded patients / samples and inconclusive results

Fig. 2. Recommendations for the reporting of genetic diagnostic test accuracy for RhD NIPD.

of study findings. Randen et al. [34] reported a decrease in the number of false results and an increase in accuracy when only one positive result, rather than all replicate outcomes, were considered. Reported test performance is, therefore, laboratory specific and might not be replicated in routine clinical practice.

All the articles identified in this review are based on data from a single laboratory, with tests undertaken under varying protocols, and results mostly derived from small sample numbers. The establishment of large-scale, standardised RhD NIPD test protocols is a key objective of the Special Non-invasive Advances in Fetal and Neonatal Evaluation (SAFE) Network established by the European Commission ([www.safenoe.org](http://www.safenoe.org)).

To date, a major weakness in terms of the emerging evidence base of RhD NIPD has been the lack of any objective discussion regarding these various flaws and their consequences for implementation in clinical practice. Furthermore, due to the lack of peer-reviewed studies reporting on high-throughput testing of non-sensitised women, discussion of the implementation of this test application is based on results from small-scale studies. Since it implies testing a much larger number of women, the evidence base from small-scale studies should not be taken forward to large-scale testing of non-sensitised women. For instance, the number of

retests will determine the cost of the test, as well as its reliability. No articles address the issue of cost versus quality. Although our review has concentrated on the quality of studies in terms of their reporting of diagnostic accuracy, and not on scientific aspects such as the primers employed, the latter are also important. Scientific aspects will inevitably affect the ultimate accuracy of the test in the wider population since they determine which of the many variants would be detected or missed.

In conclusion, based on our findings we would recommend that following standardisation to identify best practice for methodology and materials for RhD NIPD, consensus should also be reached on the required number and handling of replicates, since test accuracy was observed to be repeatedly improved in papers by discounting dissimilar replicates [14,30,33] or sample test outcomes [2,6,14,30,35,37].

Furthermore, when carrying out meta-analyses, methods to take account of the quality of studies as well as their sample size should be considered since disregard of the quality of NIPD study design may lead to erroneous conclusions about test accuracy. High-quality trials evaluating the diagnostic accuracy of RhD NIPD tests are urgently needed. Examination of articles published to February 2008 identifies a continued absence of large-scale, peer-



reviewed trials. We would recommend that such studies adhere to the STARD checklist and the recommendations (Fig. 2) specific for NIPD, and that these are consulted at the planning stage of all future NIPD studies.

## Acknowledgement

This work is supported by the European Commission funds allocated to the SAFE Network of Excellence under the 6th Framework. Project Number: LSHB-CT-2004-503243.

## References

- [1] Lo YMD, Bowell PJ, Selinger M, et al. Prenatal determination of fetal RhD status by analysis of peripheral blood of rhesus negative mothers. *Lancet* 1993;341(8853):1147–8.
- [2] Geifman-Holtzman O, Bernstein IM, Berry SM, et al. Fetal RhD genotyping in fetal cells flow sorted from maternal blood. *Am J Obstet Gynecol* 1996;174(3):818–22.
- [3] Lo YMD, Corbetta N, Chamberlain PF, et al. Presence of fetal DNA in maternal plasma and serum. *Lancet* 1997;350:485–7.
- [4] Siva SC, Johnson SI, McCracken SA, Morris JM. Evaluation of the clinical usefulness of isolation of fetal DNA from the maternal circulation. *Aust N Z J Obstet Gynaecol* 2003;43(1):10–5.
- [5] Costa J-M, Giovangrandi Y, Ernault P, et al. Fetal RhD genotyping in maternal serum during the first trimester of pregnancy. *Br J Haematol* 2002;119:255–60.
- [6] Legler T, Lynen R, Maas JH, et al. Prediction of fetal RhD and Rh CcEe phenotype from maternal plasma with real-time PCR. *Transfus Apher Sci* 2002;27(3):217–23.
- [7] Zhong XY, Holzgreve W, Hahn S. Detection of fetal Rhesus D and sex using fetal DNA from maternal plasma by multiplex polymerase chain reaction. *BJOG* 2000;107(6):766–9.
- [8] Cunningham J, Yates Z, Hamlington J, Mason G, Mueller R, Miller D. Non-invasive RNA-based determination of fetal Rhesus D type: a prospective study based on 96 pregnancies. *BJOG* 1999;106:1023–8.
- [9] Finning KM, Martin PG, Soothill PW, Avent ND. Prediction of fetal D status from maternal plasma: introduction of a new noninvasive fetal RhD genotyping service. *Transfusion* 2002;42:1079–85.
- [10] Chilcott J, Tappenden P, Jones ML, et al. The economics of routine antenatal anti-D prophylaxis for pregnant women who are rhesus negative. *BJOG* 2004;111:903–7.
- [11] Aubin J-T, Le Van Kim C, Mouro I, et al. Specificity and sensitivity of RHD genotyping methods by PCR-based DNA amplification. *Br J Haematol* 1997;98:356–64.
- [12] Daniels G, Finning K, Martin P, Soothill P. Fetal blood group genotyping from DNA from maternal plasma: an important advance in the management and prevention of haemolytic disease of the fetus and newborn. *Vox Sang* 2004;87:225–32.
- [13] Machado IN, Castillo L, Pellegrino Jr J, Barini R. Fetal RhD genotyping from maternal plasma in a population with a highly diverse ethnic background. *Rev Assoc Med Bras* 2006;52(4):232–5.
- [14] Brojer E, Zupanska B, Orzinska A, Kalinska A. Noninvasive determination of fetal RHD status by examination of cell-free DNA in maternal plasma. *Transfusion* 2005;45:1473–80.
- [15] Gautier E, Benachi A, Giovangrandi Y, et al. Fetal RhD genotyping by maternal serum analysis: a two year experience. *Am J Obstet Gynecol* 2005;192:666–9.
- [16] Finning KM, Martin P, Daniels G. A clinical service in the UK to predict fetal Rh (Rhesus) D blood group using free fetal DNA in maternal plasma. *Ann N Y Acad Sci* 2004;1022:119–23.
- [17] Van der Schoot E, Bleker O, Bonsel G, Christiaens G, Kanhai H. ZonMw. Final Report Health Care Efficiency Research Programme on <http://www.zonmw.nl>; 2005.
- [18] Szczepura A, Osipenko L, Freeman K. Non-invasive prenatal diagnosis (NIPD): Fetal RHD Genotype: Early-HTA Report. SAFE NoE; 2007, Available at: [www.safenoe.org](http://www.safenoe.org).
- [19] Bianchi DW, Avent ND, Costa J-M, van der Schoot CE. Noninvasive prenatal diagnosis of fetal Rhesus D. Ready for Primer Time. *Obstet Gynecol* 2005;106(4):841–4.
- [20] Geifman-Holtzman O, Grotegut CA, Gaughan JP. Diagnostic accuracy of non-invasive fetal Rh genotyping from maternal blood—a meta-analysis. *Am J Obstet Gynecol* 2006;195(4):1163–73.
- [21] Bossuyt MP, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Clin Chem* 2003;49(1):1–6.
- [22] Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem* 2003;49(1):7–18.
- [23] Zhong XY, Holzgreve W, Hahn S. Risk free simultaneous prenatal identification of fetal Rhesus D status and sex by multiplex real-time PCR using cell free fetal DNA in maternal plasma. *Swiss Med Wkly* 2001;131(5–6):70–4.
- [24] Lo YMD, Hjelm NM, Fidler C, et al. Prenatal diagnosis of fetal RhD status by molecular analysis of maternal plasma. *N Engl J Med* 1998;339:1734–8.
- [25] Rouillac-Le Sciellour C, Puillandre P, Gillot R, et al. Large-scale pre-diagnosis study of fetal RHD genotyping by PCR on plasma DNA from RhD-negative pregnant women. *Mol Diagn* 2004;8(1):23–31.
- [26] Gonzalez-Gonzalez C, Garcia-Hoyos M, Trujillo-Tiebas MJ, et al. Application of fetal DNA detection in maternal plasma: a prenatal diagnosis unit experience. *J Histochem Cytochem* 2005;53(3):307–14.
- [27] Lo YM, Lau TK, Zhang J, et al. Increased fetal DNA concentrations in the plasma of pregnant women carrying fetuses with trisomy 21. *Clin Chem* 1999;45:1747–51.
- [28] Hromadnikova I, Vesela K, Benesova B, et al. Non-invasive fetal RHD and RHCE genotyping from maternal plasma in alloimmunized pregnancies. *Prenat Diagn* 2005;25(12):1079–83.
- [29] Bischoff FZ, Nguyen DD, Maequez-Do D, Moise Jr KJ, Simpson JL, Elias S. Noninvasive determination of fetal RhD status using fetal DNA in maternal serum and PCR. *J Soc Gynecol Investig* 1999;6:64–9.
- [30] Hromadnikova I, Vechetova L, Vesela K, et al. Non-invasive fetal RhD exon 7 and exon 10 genotyping using real-time PDR testing of fetal DNA in maternal plasma. *Fetal Diagn Ther* 2005;20(4):275–80.
- [31] Zhou L, Thorson JA, Nugent C, Davenport RD, Butch SH, Judd WJ. Noninvasive prenatal RHD genotyping by real-time polymerase chain reaction using plasma from D-negative pregnant women. *Am J Obstet Gynecol* 2005;193(6):1966–71.
- [32] Minion JM, Schaaps JP, Retz MC, Dricot JF, Foidart JM, Senterre JM. Utilisation en routine clinique du genotypage foetal RHD sur plasma maternel: bilan de deux ans d'activite. *J Gynecol Obstet Biol Reprod* 2005;34:448–53.
- [33] Hromadnikova I, Vechetova L, Vesela K, Benesova B, Doucha J, Vlk R. Non-invasive Fetal RhD and RhCE Genotyping using real-time PCR testing of maternal plasma in RhD negative pregnancies. *J Histochem Cytochem* 2005;53(3):301–5.
- [34] Randen I, Hauge R, Kjeldsen-Kragh J, Fagerhol MK. Prenatal genotyping of RHD and SGY using maternal blood. *Vox Sang* 2003;85:300–6.
- [35] Clausen FB, Krog GR, Rieneck K, et al. Reliable test for prenatal prediction of fetal RhD type using maternal plasma from RhD negative women. *Prenat Diagn* 2005;25(11):1040–4.
- [36] Guz K, Brojer E, Zupanska B, Orzinska A, Kalinska A, Bec JR. Non-invasive fetal RhD typing and RhD negative pregnant women—preliminary observations. *Obs Ginek Pol* 2004;75:21–5.
- [37] Nelson M, Eagle C, Langshaw M, Popp H, Kronenberg H. Genotyping fetal DNA by non-invasive means: extraction from maternal plasma. *Vox Sang* 2001;80:112–6.
- [38] Al-Mufti R, Howard C, Overton T, et al. Detection of fetal messenger ribonucleic acid in maternal blood to determine fetal RhD status as a strategy for non-invasive prenatal diagnosis. *Am J Obstet Gynecol* 1998;179(1):210–4.
- [39] Turner MJ, Martin CM, O'Leary JJ. Detection of fetal Rhesus D gene in a whole blood of women booking for routine antenatal care. *Eur J Obstet Gynecol Reprod Biol* 2003;108:29–32.
- [40] Leeflang M, McDonald S, Scholten RJPM, Reitsma H, Rutjes AWS. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE (Protocol). *Cochrane Database of Systematic Reviews* 2007, Issue 2. Art. No.: MR000022. Doi: 10.1002/14651858.MR000022.pub2.
- [41] Moise Jr KJ. Noninvasive prenatal RhD testing. *Am J Obstet Gynecol* 2006;195:e20.
- [42] Roque H. Fetal RhD genotyping by maternal serum analysis: a two-year experience. *Am J Obstet Gynecol* 2006;194:905–6.
- [43] Van der Schoot CE, Ait Soussan A, Koelewijn J, Bonsel G, Paget-Christiaens LGC, de Haas M. Non-invasive antenatal RHD typing. *Transfus Clin Biol* 2006;13:53–7.